Role of Modular Polyketide Synthases in the Production of Polyether Ladder Compounds in Ciguatoxin-producing Gambierdiscus polynesiensis and G.excentricus (Dinophyceae)

DR. GURJEET S KOHLI (Orcid ID : 0000-0002-4578-2355)

Received Date : 06-Apr-2016

Revised Date : 31-Jan-2017

Accepted Date : 03-Feb-2017

Article type : Original Article

Role of Modular Polyketide Synthases in the Production of Polyether Ladder Compounds in Ciguatoxin-producing *Gambierdiscus polynesiensis* and *G.excentricus* (Dinophyceae)

Gurjeet S. Kohli[a, b], Katrina Campbell[c], Uwe John[d], Kirsty F. Smith[e], Santiago Fraga[f], Lesley L. Rhodes[e], Shauna A. Murray[a]

[a] Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Ultimo, NSW 2007, Australia

[b] Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 689528

[c] Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast, UK BT9 5AG

[d] Alfred Wegener Institute for Polar and Marine Research, Bremerhaven 27515, Germany

[e] Cawthron Institute, 98 Halifax Street East, Nelson 7010, New Zealand

[f] Instituto Español de Oceanografía, Centro Oceanográfico de Vigo, Subida a Radio Faro 50, 36390 Vigo, Spain

Corresponding author:

Shauna A. Murray & Gurjeet S. Kohli; E-mail: shauna.murray@uts.edu.au gurjeet.kohli@uts.edu.au; gurukohli@gmail.com Tel: +61 2 9514 8404; Fax: +61 2 9514 4079; CB07.06.039, PO Box 123 Broadway, NSW 2007, Australia

Abstract

Gambierdiscus, a benthic dinoflagellate, produces ciguatoxins that cause the human illness Ciguatera. Ciguatoxins are polyether ladder compounds that have a polyketide origin, indicating that polyketide synthases (PKS) are involved in their production. We sequenced transcriptomes of *G. excentricus* and *G. polynesiensis* and found 264 contigs encoding single domain ketoacyl synthases (KS; *G. excentricus*: 106, *G. polynesiensis*: 143) and ketoreductases (KR; *G. excentricus*: 7, *G. polynesiensis*: 8) with sequence similarity to type I PKSs, as reported in other dinoflagellates. Additionally, 24 contigs (*G. excentricus*: 3, *G. polynesiensis*: 21) encoding multiple PKS domains (forming typical type I PKSs modules) were found. The proposed structure produced by one of these megasynthases resembles a partial carbon backbone of a polyether ladder compound. Seventeen contigs encoding single domain KS, KR, s-malonyltransacylase, dehydratase and enoyl-reductase with sequence similarity to type II fatty acid synthases (FAS) in plants were found. Type I PKS and type II FAS genes were distinguished based on the arrangement of domains on the contigs and their sequence similarity and phylogenetic clustering with known PKS/FAS genes in other organisms. This differentiation of PKS and FAS pathways in *Gambierdiscus* is important, as it will facilitate approaches to investigating toxin biosynthesis pathways in dinoflagellates.

Keywords:

Ciguatoxin, *Gambierdiscus*, biosynthesis, polyketide synthase, fatty acid synthase

CIGUATERA Fish Poisoning (CFP) is a syndrome caused by the consumption of seafood contaminated with analogues of the toxin ciguatoxin (CTX). It is common in tropical countries worldwide, and impacts ~50,000-500,000 people annually, despite being significantly underreported (Fleming et al. 1998, Friedman et al. 2008). In Pacific Small Island Developing States, The rate of CFP cases reported has increased recently by 60% over a 10-year period (Skinner et al. 2011).

Chemically, CTXs are thermostable and liposoluble cyclic polyether ladder compounds, with more than 11 congeners now described (Kohli et al. 2015a and references therein). These congeners can be divided into three types based on structural changes, geographical origin and toxicity (P-CTXs from the Pacific, C-CTXs from the Caribbean/Atlantic and I-CTXs from the Indian Oceans) (Kohli et al. 2015a and references therein). In the Pacific, P-CTX-1B is the principle toxin present in carnivorous fish implicated in CFP (Murata et al. 1990, Lewis et al. 1991). Though no strains of *Gambierdiscus* have been shown to produce P-CTX-1B, there is evidence to support that P-CTX-1B, P-CTX-2 and P-CTX-3 are derived from the biotransformation of CTXs produced by *Gambierdiscus*, P-CTX-4A and P-CTX4B (Murata et al. 1990, Lewis and Holmes 1993, Yasumoto et al. 2000). *G. polynesiensis*, first described from French Polynesia in the Pacific Ocean region (Chinain et al. 1999) produces both Type 1 (CTX-4A, CTX-4B) and Type 2 (CTX-3C, M-seco-CTX-3C, 49-epiCTX-3C) P-CTXs (Chinain et al. 2010, Rhodes et al. 2014) in culture. Multiple strains of this species have now been studied for CTX

production, and non-producing strains of *G. polynesiensis* have not yet been reported (Chinain et al. 2010, Rhodes et al. 2014, Murray et al unpub data). *G. excentricus*, a recently described species from Canary Islands in the Atlantic Ocean region (Fraga et al. 2011), was found to produce CTX based on the Neuro-2a cell based assay, though its toxin profile has not yet been characterised (Fraga et al. 2011).

The proposed mechanism of toxicity of CTXs is based on its ability to activate sodium channels in cells with excitable membranes such as the peripheral nerve cells. When CTXs binds to the sodium channels, there is a massive influx of sodium ions, resulting in the depolarisation of the cells and onset of spontaneous action potentials in effected cells (Kohli et al. 2015a and references therein). This causes various gastrointestinal and neurological symptoms in humans (Kohli et al. 2015a and references therein).

CTX toxin profiles and structures have been determined by liquid chromatography-mass spectrometry (LCMS) techniques, accompanied by nuclear magnetic resonance (Murata et al., 1989, Murata et al., 1990a, Lewis et al., 1991, Satake et al., 1996) and radio ligand binding (Hamilton et al., 2002a, Hamilton et al., 2002b). These methods require purified CTX standards. Confirmation of toxin by LCMS involves the isolation and fractionation of the various CTX compounds of known molecular weights. At present, there is a very limited supply of purified CTX standards available (Berdalet et al., 2012). Currently state of the art LCMS equipment struggles to reach the low levels of detection required for CTXs, as they may be highly toxic in trace amounts. For this reason, the toxin profile of relatively few species of *Gambierdiscus* have been determined with certainty (Table 1), and of those species that have been examined, there has sometimes been a mismatch between results from cell-based assays, and those using LCMS (Table 1).

Liposoluble extracts of other *Gambierdiscus* species such as *G. australes*, *G. toxicus*, *G. pacificus* and *G. belizeanus* have been reported to be toxic via the receptor-binding assay (Chinain et al. 2010). Qualitative detection of P-CTX-3C in the liposoluble fraction of *G. belizeanus* strain CCMP401 has been reported previously (Roeder et al. 2010). However, other studies have shown that no CTXs (P-CTX-3B, P-CTX-3C, P-CTX-4A, P-CTX-4B) could be detected via LCMS in *G. belizeanus* strain CCMP401 (Kohli et al. 2015b) or in several different strains of *G. australes* (Rhodes et al. 2010, Rhodes et al. 2014, Kohli et al. 2015b). The toxicity of the liposoluble fractions of *G. australes*, *G. toxicus*, *G. pacificus* and *G. belizeanus* were low compared to that of *G. polynesiensis* in the study of strains from French Polynesia (Chinain et al. 2010). This might indicate the presence of novel congeners of CTXs and/or other polyketide compounds that inhibit sodium channels in the liposoluble extracts of these species. Therefore, detailed characterisation of the toxin profile of these species of *Gambierdiscus* via LC-MS and/or nuclear magnetic resonance is clearly required.

An understanding of the genetic basis of ciguatoxin synthesis would be immensely useful in aiding our understanding of their fundamental science, including molecular ecology, evolution, toxicology, chemistry, as well as assisting in the development of monitoring measures to protect public health.

Based on radio labelled precursor studies, and determinations of their chemical structure, it is clear that cyclic polyether compounds such as CTXs are produced by way of polyketide synthesis pathways in dinoflagellates (Kalaitzis et al. 2010, Lee et al. 1986, Chou and Shimizu 1987, Lee et al. 1989, Wright et al. 1996, Murata et al. 1998).

Fatty acid synthases (FAS) and PKSs are closely related and have a common evolutionary history (Kohli et al. 2016). The ketosynthase (KS) domain, which performs the condensation reaction between acyl units, along with the acyl transferase (AT) and acyl carrier protein (ACP) forms the core structure of FASs and PKSs (Cane et al. 1998, Khosla et al. 1999, Jenke-Kodama et al. 2005). Other domains that modify the acyl-units after condensation are dehydratase (DH), enoylreductase (ER) and ketoreductase (KR), are selectively present/absent in PKSs, however essential for FASs. The thioesterase (TE) domain hydrolyses the polyketide chain from ACP ultimately releasing the polyketide compound from the megasynthase (Cane et al. 1998, Khosla et al. 1999, Jenke-Kodama et al. 2005 and references therein).

Three types of PKSs have been described so far. In iterative type I PKS, a set of catalytic domains are present in a single protein and used in a cyclic fashion repeatedly for chain elongation, analogous to fatty acid synthesis is animals and fungi (Khosla et al. 1999, Jenke-Kodama et al. 2005). Iterative type I PKSs can be further subdivided into reducing PKSs (that produce fatty acid derivatives), partially reducing PKS and non-reducing PKSs (that produce true polyketides) (Jenke-Kodama et al. 2005 and references therein). In modular type I PKSs, catalytic domains are organised in sequential modules, where each module contains all the catalytic domains needed to perform one condensation reaction and increase the length of the polyketide chain by 2 carbon atoms, until it reaches the last module that contains the TE domain, which terminates the elongation process and releases the polyketide chain (Cane et al., 1998; Jenke-Kodama et al., 2005; Khosla et al., 1999). Type II PKSs consist of multi protein complexes where each catalytic domain is on a separate peptide which function as mono-functional proteins in an iterative fashion analogous to type II FASs in bacteria and plants (Cane et al., 1998; Jenke-Kodama et al., 2005; Khosla et al., 1999). Type III PKSs are self-contained homodimeric enzymes where each monomer performs a specific function in an iterative manner without the use of acyl carrier proteins (that is, it acts directly on acyl units), also called as chalcone synthase like PKSs (Ferrer et al., 1999).

As *Gambierdiscus* species appear to possess amongst the largest genomes known from eukaryotes, (32.5 Gbp for *G. australes* and 35 Gbp for *G. belizeanus*) (Kohli et al. 2015b), similar to other dinoflagellates (Veldhuis et al. 1997, Hou and Lin 2009, Shoguchi et al. 2013, Kohli et al. 2015b), this poses a significant challenge in the ongoing effort to find the genes responsible for cyclic polyether production in dinoflagellates. Previous studies have therefore used

transcriptomic sequencing to identify candidate genes related to toxin production in *Karenia brevis* (Monroe and Van Dolah 2008), *G. polynesiensis* (Pawlowiez et al. 2014), *G. australes, G. belizeanus* (Kohli et al. 2015b), *Azadinium spinosum* (Meyer et al. 2015), *Heterocapsa circularisquama* (Salcedo et al. 2012), *Karenia mikimotoi* (Kimura et al. 2015), and *Alexandrium ostenfeldii* (Jaeckisch et al. 2011, Eichholz et al. 2012) and successfully identified hundreds of polyketide synthase (PKS) genes. Recently, PKS genes were also identified in the genome of *Symbiodinium minutum* (Beedessee et al. 2015). In dinoflagellates, contigs encoding KS domains encode one domain per transcript (a feature typically observed in type II PKSs) but phylogenetically cluster with other type I PKS-KS domains, therefore, classified as type I PKS and indicating the presence of a novel type I PKS in dinoflagellates (Monroe and Van Dolah 2008, Pawlowiez et al. 2014, Ryan et al. 2014, Kohli et al. 2015b, Kohli et al. 2016, Meyer et al. 2015, Eichholz et al. 2012). As PKSs and FASs share a similar enzymatic domain structure it is important to differentiate between the genes encoding the two groups of enzymes, in order to better understand the polyketide/toxin production in *Gambierdiscus*.

Due to several factors, including the lack of dinoflagellate model species that have been well characterised genetically, difficulties in successfully genetically manipulating dinoflagellate species, including developing knock-outs, and difficulties in axenically culturing dinoflagellates, the confirmation of genetic pathways in dinoflagellates has been difficult. To date, confirmation of genes likely responsible for certain pathways has used circumstantial information, such as the presence or absence of related genes in species with a propensity to produce that toxin (Stüken et al. 2011, Murray et al. 2015).

A comparative transcriptomic approach, in which transcriptomes of species with contrasting toxin profiles is compared, is one approach to overcome this difficulty. In this study, we generated transcriptomes of two species, *G. excentricus* (Fraga et al. 2011) and *G. polynesiensis* (Rhodes et al. 2014), which likely produced known CTX analogs, and compared them to the transcriptomes of two genetically relatively closely related, most likely non-CTX producing *Gambierdiscus* species, *G. belizeanus* and *G. australes* (Kohli et al. 2015b). The aim of this comparison was to investigate the presence and expression of candidate genes that may be linked to cyclic polyether synthesis in producing and non-producing similar species.

MATERIAL AND METHODS

Culture conditions, RNA extraction and assembly

Gambierdiscus polynesiensis (CAWD212) was isolated from Rarotonga, Cook Islands (21°13'24.18" S; 15°43'54.94" W) in March 2013 (Rhodes et al. 2014). *Gambierdiscus excentricus* (VGO790) was collected on March 28th, 2004 as an epiphyte on small filamentous macroalgae and turf on a tidal pond in Punta Hidalgo 28º 34'27''N, 16º 19'52''W, Tenerife Island, Spain (Fraga et al. 2011). Strain VGO790 is barcoded in GenBank (GenBank ID: JF303074, large ribosomal subunit, D8-D10, 771bp), (GenBank ID: HQ877874, large ribosomal

subunit, D1-D3, 822bp) and (GenBank ID: JF303065, large ribosomal subunit, D1-D3, 822bp) (Fraga et al. 2011). Strains were grown at 25 °C under cool white fluorescent light (light intensity 60 µmol m^{-2} s^{-1}; 12:12 light:dark cycle) in in f/2 medium (Guillard and Ryther 1962).

RNA extraction and library preparation was carried out as described previously (Kohli et al. 2015b). The *G. excentricus* cDNA library was sequenced on a half lane of HiSeq2000 (Illumina, San Diego, CA), generating a total of 135,064,950 100 bp paired-end reads. The *G. polynesiensis* cDNA library was sequenced using two lanes of HiSeq2000 (Illumina, San Diego, CA), generating a total of total of 1,064,432,504 100 bp paired-end reads. Raw reads were quality filtered and assembled into contigs using CLC Genomics Workbench Version 7.0 (CLC bio, Cambridge, MA) using default software settings and a total of 883,662,391 (*G. polynesiensis*) and 118,044,360 (*G. excentricus*) reads mapped during assembly. During the assembly, scaffolding was performed and the assembly was also validated using read mapping. Any contigs less than 300bp of length and/or containing gaps (NNNNs) of more that 100bp were not analysed any further.

Transcriptome Analysis

BLASTx analysis was performed at an E-value cutoff of 10^{-9} against the nr database of GenBank. BLAST2GO (Götz et al. 2008) and Kyoto Encyclopedia of Gene and Genomes (KEGG) analysis using default software settings were used to perform mapping and annotation of transcripts. To improve the annotation of the transcriptomes, InterProScan analysis was carried out in BLAST2GO which uses the following databases: BlasProDom, FPrintScan, HMMPIR, HMMPfam, HMMSmart, HMMTigr, ProfileScan, HAMAP, PatterScan, SuperFamily, SignalPHMM, TMHMM, HMMPanther, Gene3D, Phobius, Coils (Mitchell et al. 2014). To analyse the comprehensiveness of the transcriptomes, the Core Eukaryotic Genes Mapping approach (CEGMA) tool (Parra et al. 2007) was also used. Initial search for PKS/FAS genes was carried out by text searching the annotations (ketosynthase, PKS, polyketide synthase, ketoreductases, FAS). PKS/FAS genes were also identified using HMMER (Finn et al. 2011), in which in-house HMM databases were developed using sequences identified via text searching the annotations and already known sequences from other dinoflagellates for each enzyme investigated in this study. Alignments used for the HMM databases have been provided in supplementary file 1. After obtaining the sequences from HMMER, Pfam (Punta et al. 2012) and CDD (Marchler-Bauer et al. 2014) were used for identification of conserved amino acid residues and functional prediction of PKS/NRPS and FAS genes. Sequences identified by HMMER/BLASTx where no functional predictions of PKS/NRPS and FAS genes using Pfam and CDD could be obtained, were discarded and not analysed any further. Identification of transit peptide targeted towards chloroplast was detected using ChloroP (Emanuelsson et al. 1999). Contigs containing the sequence "DCCGTAGCCATTTTGGCTCAAG (D = T, A, or G)" within the first/last 100bp were considered to have a full spliced leader (all 22 bp)/partial spliced leader (11-21 bp,) sequence. Sequence clustering and comparison was carried out on web server CD-HIT (Huang et al. 2010).

Raw data was submitted to the NCBI sequence read archive under accession numbers SRR3348983 (*G. excentricus*) and SRR3358210 (*G. polynesiensis*). Full-length PKS and FAS genes were submitted to Genbank under accession numbers KX395751 - KX395902 and the rest of the assembled sequences (combined with partial PKS/FAS sequences) were submitted to the NCBI transcriptome shotgun assembly archive (*G. excentricus* GETL00000000; *G. polynesiensis* GETK00000000) under Bioproject numbers PRJNA317708 (*G. excentricus*) and PRJNA317942 (*G. polynesiensis*).

Phylogenetic Analysis

MAFFT (Katoh et al. 2002) and/or Clustal W (Thompson et al. 1994) aligners within Geneious® software (Kearse et al. 2012) were used for sequence alignments. Alignments were manually inspected and trimmed to ensure they spanned the same coding region. Phylogenetic analysis was carried out in RAxML Version 7.0 (Stamatakis 2006) using GAMMA and LG models of rate heterogeneity with 1000 bootstraps.

RESULTS AND DISCUSSION

Transcriptomic analysis

Comprehensive transcriptomic libraries of *G. excentricus* and *G. polynesiensis* were analysed, and resulted in 77,393 (Mean length: 1148, GC content: 62.3%) and 115,780 (Mean length: 924.5, GC content 61.6%) contigs > 300 bp, respectively. The GC content of the transcriptomes was high (62.3 % for *G. excentricus* and 61.6 % for *G. polynesiensis*) as compared to other eukaryotes (Shoguchi et al. 2013, Pawlowiez et al. 2014, Kohli et al. 2015b, Meyer et al. 2015, Ryan et al. 2014). For *G. excentricus,* it was determined using BLASTx analysis that at an e-value cut-off of 10^{-9}, 24.7% of the contigs had an annotated match, 17.4 % had a non-annotated match and 57.9% of the contigs lacked similarity to any sequences in the nr database of GenBank (Table 2). For *G. polynesiensis* it was determined using BLASTx analysis that at an e-value cut-off of 10^{-9}, 20.9% of the contigs had an annotated match, 16.1 % had a non-annotated match and 63% of the contigs lacked similarity to any sequences in the nr database of GenBank. This is in the range for other protist studies (Keeling et al. 2014) but particular for dinoflagellates (Pawlowiez et al. 2014, Ryan et al. 2014, Kohli et al. 2015b, Meyer et al. 2015). Read mapping/coverage analysis of *G. excentricus* and *G. polynesiensis* transcriptomes showed that 72.1% and 88.4% of contigs, respectively, had greater than 20x coverage (Table 2). Excluding the lower coverage sequence did not affect the high difference in contig numbers between both species and therefore we kept all contigs in the analyses. CD-HIT cluster analysis was performed (at 90%, 95% and 98% sequence similarity at nucleotide level) to identify highly similar sequences within and between both the transcriptomes to identify potential isoforms and false positive transcript candidates. The analysis revealed that 98.6-99.9% of contigs in *G. excentricus* (at different sequence similarity thresholds) and 90.1- 99.3% of contigs in *G. polynesiensis* were unique and this indicates that these contigs are not just isoforms from the same gene in the

respective datasets. To compare the two transcriptomes a BLASTn analysis (e-value cut-off of 10^{-20}) was performed, where *G. polynesiensis* contigs were used as a query against a database of *G. excentricus* contigs. The analysis revealed that 66.8 % of *G. polynesiensis* contigs had a positive match to the *G. excentricus* transcriptomic database. However, the percentage similarity of 93.6 % of the *G. polynesiensis* sequences that matched the *G. excentricus* transcriptome was 70-90%, indicating that the two species are genetically quite distant to each other. This is also evident in the 18S ribosomal phylogeny of *Gambierdiscus* and *Alexandrium* species (Figure S1), in which *Gambierdiscus* species occur on much longer branch lengths as compared to *Alexandrium* species (Figure S1). Similar BLASTn comparisons were carried out among the transcriptomic datasets of *G. excentricus*, *G. belieanus*, *G. australes* and *G. polynesiensis*. The analysis revealed that 63-75% contigs of the query species had positive matches in the transcriptome of the species used as a database.

In the *G. excentricus* transcriptome, 8890 (11.5 % of the transcriptome) contigs had a full or partial (defined as >11bp within 100bp of ends of contig) dinoflagellate-specific spliced leader sequence (SL) (Lidie and Van Dolah 2007), at the 5' end and Similarly, in the *G. polynesiensis* transcriptome, 6869 (6% of the transcriptome) contigs had a full or partial SL sequence.

As reported in studies of the genomes or transcriptomes of *Symbiodinium kawagutii* (Lin et al. 2015), *Symbiodinium minutum* (Bayer et al. 2012, Shoguchi et al. 2013), *Lingulodinium polyedrum* (Roy and Morse 2012), *Azadinium spinosum* (Meyer et al. 2015) and two *Gambierdiscus* species (Kohli et al. 2015b), contigs encoding a full suite of essential histone-coding proteins (H2A, H2B, H3, H4) were found in the transciptomes of *G. excentricus* and *G. polynesiensis* (Table S1). Transcriptomes of *G. excentricus* and *G. polynesiensis* also encoded 95% (*G. excentricus*: 159 of 167 and *G. polynesiensis*: 161 of 167) of the essential enzymes required for C-3 carbon cycle, oxidative phosphorylation, pentose phosphate pathway, glycolysis, tricaroxalic acid cycle, purine nucleotide synthesis (Inosine monophosphate synthesis; synthesis of AMP, ADP, ATP, synthesis of dADP, dATP; synthesis of GMP, GDP, GTP and synthesis of dGDP and dGTP), pyrimidine nucleotide synthesis (Uridine monophosphate synthesis, synthesis of UDP, UTP, CTP and CDP; synthesis of dCDP, dCTP, dUDP and dUTP; synthesis of dTMP, dTDP and dTTP), tyrosine-phenylalanine-tryptopham synthesis, serine-glycine-threonien synthesis, arginine-proline synthesis, Alanine-Aspartic acid-Asparginine-Glutamic acid-Glutamine synthesis, Cysteine-Methionine synthesis, valine-leucine-isoleucine synthesis, lysine synthesis and histidine synthesis (Table S2).

To further assess the completeness of the transcriptomes, we used the core eukaryotic genes mapping approach (CEGMA) (Parra et al. 2007), and found that the transcriptomes contained 78 % (361 of 458) and 74.6 % (342 of 458) of the core eukaryotic genes for *G.excentricus* and *G. polynesiensis*, respectively. This is comparable to other dinoflagellate transcriptomes investigated via CEGMA analysis to date (Ryan et al. 2014, Kohli et al. 2015b, Meyer et al. 2015).

Fatty acid and polyketide biosynthesis in *Gambierdiscus*

Fatty acid synthesis

Plants have type II FASs, which carry each catalytic domain on separate polypeptides that form multi-protein complexes (Jenke-Kodama et al. 2005 and references therein). In plants, a separate nuclear gene encodes each polypeptide that is targeted towards the chloroplast, where fatty acid synthesis occurs (McFadden, 1999 and references therein). Gene knockout studies and functional characterisation of all the FAS catalytic domains has been carried out in higher plants (White et al., 2005; Brown et al., 2010 and references therein). Recently, these type II FAS genes resembling that of plants, were reported in many dinoflagellates and other protists (116 genera) (Kohli et al., 2016).

We found genes encoding 3-ketoacyl ACP synthase I, II & III (KASI-FabB, KASII-FabF, KASIII-FabH), ACP s-malonyltransacylase (AT-FabD), trans3-ketoacyl ACP reductase (KR- FabG), 3-hydroxyacyl-ACP dehydratase (DH-FabZ), and enoyl-ACP reductase (ER-FabI) in the transcriptomes of *G. australes, G. belizeanus, G. excentricus and G. polynesiensis* (Table S3) that are likely involved in fatty acid synthesis. As seen in other dinoflagellates and protists (Kohli et al. 2016), each transcript encoded individual FAS enzyme/domain (a feature typically observed in type II FASs) in *Gambierdiscus*. Phylogenetic analysis of 32 type II 3-ketoacyl ACP synthase II and 71 type I ketosynthase domains from prokaryotic and eukaryotic PKS and FASs, revealed that *Gambierdiscus* 3-ketoacyl ACP synthase II genes cluster with other type II FAS genes from phototrophic lineages such as green algae and plants (Figure 1), indicating the presence of a type II FAS in *Gambierdiscus*. Unlike other protists, only one type of gene family encoding KASII-FabF, which carries out the condensation reaction during elongation of the fatty acid chain, was found in *Gambierdiscus*. Transit peptides targeted towards the chloroplast were found in 70% of the contigs encoding type II FAS enzymes (Table S3). Residues comprising the active sites for the seven enzyme involved in type II FAS were confirmed in all four species of *Gambierdiscus* (Table S3). These features indicate that the genes are fully functional, and that whole or part of the process of fatty acid synthesis appears to take place in the chloroplast.

In order to understand the evolutionary history of these genes, a concatenated phylogeny of 5 type II FAS enzymes (KASIII-FabH, AT-FabD, DH-FabZ, ER ER-FabI, KR- FabG) was carried out. *Chromera velia* was used as an out-group, and dinoflagellates formed a distinct well-supported monophyletic clade (Figure 2). Within the dinoflagellate clade, the evolution of these genes broadly follows the trend of dinoflagellate evolution, in which the orders Gonyaulacales, Peridiniales (except *Azadinium*) and Suessiales form monophyletic clades. All four species of *Gambierdiscus* group together within the Gonyaulacales clade, and *Alexandrium* formed the sister clade, indicating that it is evolutionarily closely related to *Gambierdiscus* (Orr et al. 2012, Murray et al. 2015).

Polyketide biosynthesis

Since KS domains are an essential and highly conserved domain, involved in PKS biosynthesis (Kroken et al. 2003, John et al. 2008), the search for genes encoding this domain was the primary focus of this study. In *Gambierdiscus*, other studies have reported KS domains from *G. australes* and *G. belizeanus* (*G australes*: 90 full and 12 partial; *G. belizeanus*: 74 full and 40 partial) (Kohli et al. 2015b) and 22 KS domains in *G. polynesiensis* strain TB-92 (Pawlowiez et al. 2014). As these contigs coding KS domains, encoded one domain per transcript (a feature typically observed in type II PKSs) but phylogenetically clustered with other type I PKS-KS domains, they were classified as type I PKSs (Pawlowiez et al. 2014, Kohli et al. 2015b).

In this study, contigs encoding KS domains were found to be highly numerous and diverse in the transcriptomes of *G. excentricus* and *G. polynesiensis*. In *G. excentricus* transcriptome, contigs encoded 86 full and 20 partial KS domains and 7 full KR domains were found (Table 3, Table S4). In *G. polynesiensis* transcriptome, contigs encoded 73 full and 70 partial KS domains and 1 partial and 7 full KR domains were found (Table 3, Table S4). Similar to other dinoflagellates, these contigs encoding KS domains in *G. excentricus* and *G. polynesiensis* also encoded one domain per transcript (Table 3). Dinoflagellate transcripts encoding single KS domains have a centrally conserved ExExGYLG motif at the N-terminal region (John et al., 2012). Eighteen transcripts encoding a single KS domain in each of the *G. polynesiensis* and *G. excentricus* transcritpomes also encoded the conserved N-terminal motif ExExGYLG.

In addition to the transcripts encoding single PKS (KS) domains, 24 contigs encoding multiple PKS domains per transcript were also found in *G. excentricus* and *G. polynesiensis* (Table 3). None of these transcripts encoded a dinoflagellate specific SL at the 5' end of the sequence. In previous studies, due to the lack of a dinoflagellate specific SL and clustering of these KS domains within the bacterial type I PKS clade during phylogenetic analysis, these contigs (encoding multiple PKS domains) were considered likely to have been part of the transcriptomes of a bacterial contaminant of the strain, as the strains were not axenic (Kohli et al. 2015b). However, recently a transcript encoding a hybrid polyketide synthase – nonribosomal peptide synthase (PKS-NRPS) was found in *Amphidinium carterae* (Bachvaroff et al. 2015). This gene was found to encode dinoflagellate specific SL and a polyA tail, indicating its dinoflagellate origin, despite the fact that it clustered within the bacterial type I PKS clade based on phylogenetic analysis (Bachvaroff et al. 2015). A transcript encoding partial hybrid PKS-NRPS was also found in *G. polynesiensis* transcriptome (contig 19937, Table 3). These results indicate that dinoflagellates can express genes incorporated into the genome of apparently relatively recent bacterial origin, likely through processes such as lateral gene transfer, which have been found to be likely responsible for the novel introduction of some of the genes for saxitoxin synthesis into dinoflagellates species (Stüken et al. 2011). This indicates that studies attempting to identify genes involved in dinoflagellate secondary metabolite biosynthetic pathways need to consider both genes of clear dinoflagellate origin and those with relatively recent bacterial homology, as well as to consider that multiple gene duplications, loss and selection for novel functions are also likely to have occurred (Murray et al. 2015).

Phylogenetic analysis of all the KS domains found in this study, and from *G. australes* and *G. belizeanus* (Kohli et al. 2015b) (Table 3), was performed to examine the evolutionary history of the KS domains and identify potential bacterial contaminants (based on the phylogenetic position of transcripts). Transcripts encoding single KS domains grouped with other protistan type I PKS sequences with high bootstrap support and separate from other prokaryotic PKS sequences (Figure 3). Within the protistan clade, transcripts encoding single KS domains from dinoflagellates formed a distinct well-supported clade as seen in previous studies (Monroe and Van Dolah 2008, Eichholz et al. 2012, Pawlowiez et al. 2014, Kohli et al. 2016). Within the dinoflagellate clade of transcripts encoding single KS domains, they formed three sub-clades (Clade A, B and C), as reported previously (Kohli et al. 2016). Clade A and C consisted of KS domains with all their active site residues intact (Figure 3). KS domains in clade B did not have one or more active site residues intact (Figure 3). KS domains from contigs containing multiple domains formed six separate clades (Figure 3). Clade D consisted of KS domains from: modular type I PKS (cis AT) from bacteria and the recently discovered hybrid PKS-NRPS found in *Amphidinium, Heterocapsa* and *G. polynesiensis* (Figure 3). Clade D also contained multiple other KS domains from all four *Gambierdiscus* species, however as the transcripts were incomplete it could not be determined if these contigs were PKS-NRPS hybrids or typical type I PKSs. Chlorophytes, apicomplexa and haptophytes are also known to possess typical modular type I PKSs. In addition to the sub-clades consisting of dinoflagellate monofunctional and multiple KS domains (Figure 3), the protistan clade also consisted of well-supported and distinct sub-clades of KS domains from chlorophytes, apicomplexa and haptophytes (Figure 3).

G. australes and *G. belizeanus* produce MTXs (Table 1, *G. australes* : MTX-1 and putative MTX-3; *G. belizeanus* : putative MTX-3), and the strains used in this study have not been found to produce P-CTX-3B, P-CTX-3C, P-CTX-4A, P-CTX-4B using LC-MS analysis (Kohli et al. 2015b, Rhodes et al. 2010). *G. polynesiensis* is a potent producer of CTXs, however does not produce MTX-1, but only putative-MTX-3 (Rhodes et al. 2014). Based on toxicity assays, *G. excentricus* appears to produce both CTXs and MTXs (Fraga et al. 2011). Despite the fact that this group of species of *Gambierdiscus* produce different types of toxins, no differences were found in the expression of monofunctional KS domains in the four species. We observed 12 groups of monofunctional KS domains in clade C and A, which were 98-99 % similar to each other at amino acid level and 88-93 % similar at the nucleotide level (Figure 3). It is possible that some of these genes may be involved in the production of a polyketide compound produced by all strains. This compound might be a new congener of CTX/MTX or any other type of polyether/polyketide compound. Other species/strains of *Gambierdiscus* are known to produce a range of other polyketide compounds such as gambieric acids (Nagai et al. 1992), gambierol (Satake et al. 1993) and gambieroxide (Watanabe et al. 2013). Therefore it is imperative to study the toxin/polyketide profile of specific strains of *Gambierdiscus* in order to determine the biosynthesis genes responsible for producing these specific compounds.

The discovery of sequences encoding typical modular type I PKSs in *Gambierdiscus,* which may be of dinoflagellate origin*,* opens new avenues of investigation (Table 3). Contig 50464 found in *G. polynesiensis* encoded a partial typical type I modular PKS that consisted of 6 modules (Length: 30289 bp; Open read frame: 30141 bp; GC content: 57.9 %). KS domains encoded in

this contig clustered within the protistan clade (Figure 3). The last module consisted of TE domain, which indicates that it is the last module in the megasynthase. A structure that could be produced by this PKS has been proposed, which resembles a partial carbon backbone of a polyether ladder compound (Figure 4). Sequences resembling *G. polynesiensis* contig 50464 were also found in *G. excentricus* (contig 6057; encodes: DH-ER-KR-ACP-TE; sequence similarity: 85.2% at amino acid level), *G. australes* (contig 20703; encodes: ACP-KS-AT-DH-ER-KR-ACP-TE; sequence similarity: 84.2% at amino acid level) and *G. belizeanus* (contig 11718; encodes: AT-DH-ER-KR-ACP-TE; sequence similarity: 81.8% at amino acid level).

In dinoflagellates, the origin of transcripts encoding single type I PKS catalytic domains has been attributed to a secondary separation of typical type I multimodular PKS into single domains/enzymes (Monroe and Van Dolah 2008; Eichholz et al. 2012). This indicates that these single PKS enzymes/domains might work iteratively in multiple PKS pathways to produce different polyketides. However, this study has also revealed the presence of typical type I multimodular PKS such as contig 50464 in dinoflagellates that possess large gene clusters which encode all the enzymes/domains in a sequential manner (open reading frame) required for polyketide biosynthesis. Different evolutionary and functional constraints might explain why some PKSs retain the modular gene cluster structure and others are separated into genes encoding single enzymes/domains; often exhibiting high sequence diversity within KS genes of one species (Kohli et al., 2016).

No gene clusters associated with the biosynthesis of polyether ladder compounds have been elucidated from a dinoflagellate to date. Monensin and nanchangmycin are non-ladder polyether compounds, structurally closest to polyether ladders, and are produced by modular type I PKSs (Leadlay et al. 2001, Sun et al. 2003). In monensins biosynthesis, the carbon backbone is produced via a modular type I PKS assembly and undergoes epoxidation and polyepoxide cyclisation that might be performed by epoxidases and epoxide hydrolases that are also encoded in the monensin biosynthesis gene cluster (Leadlay et al. 2001, Oliynyk et al. 2003). Similar to the proposed biosynthesis pathway for brevetoxins (BTX) (Lee et al. 1989, Rein and Snyder 2006) and MTX-1 (Kohli et al. 2015b), we propose a possible CTX biosynthesis pathway in which the carbon backbone is produced via polyketide biosynthesis, followed by epoxidation and polyepoxide cyclisation carried out via PKSs, epoxidases and epoxide hydrolases (Figure 5). In support of the above proposed biosynthesis pathway we also found sequences that encoded epoxidases/monooxygenases and epoxide hydrolases in *G. excentricus* and *G. polynesiensis* (Table S5). These sequences encoded 13 different types of epoxidases/monooxygenases and 4 different types of epoxide hydrolases in *G. excentricus* and *G. polynesiensis*. In contrast to an earlier theory stipulating epoxidation and cyclisation of polyether ladders during polyketide synthesis (Shimizu 2003), the hypothesis here proposes that these processes might occur after the carbon backbone is synthesised by PKSs.

CONCLUSION

In conclusion, the gene catalogues of *G. polynesiensis* and *G. excentricus* presented here are amongst the most comprehensive yet found in a dinoflagellate. In addition to the genes associated with 23 regulatory pathways, the gene catalogues provide the most exhaustive libraries of full transcripts described from a single genus of dinoflagellate. In addition to the vast diversity of PKS genes, we present a clear distinction between genes responsible for fatty acid and polyketide biosynthesis in *Gambierdiscus.* We identified numerous genes related to polyketide synthesis that may be associated with toxin production in these two ciguatoxin-producing strains. The results presented here are a step towards recognising the genes which are crucial to the formation of this major group of marine biotoxins, that is currently responsible for the majority of the marine biotoxin related seafood poisonings worldwide.

ACKNOWLEDGMENTS

We thank Prof. Naresh Kumar of the University of New South Wales (Australia) for help with the proposed biosynthesis of ciguatoxin. Sequencing of the *G. excentricus* transcriptome was carried out at the Leibniz-Institute for Age Research, Fritz Lipmann Institute Jena. We thank Dr. Helena Mangs and Tonia Russell for performing transcriptomic sequencing of *G. polynesiensis* courtesy of Ramaciotti centre for genomics at University of New South Wales. We thank the Australian Research Council (FT120100704, DP120103199) for funding SM and GK. Financial support to U.J. was provided by the PACES research program of the Alfred-Wegener-Institute Helmholtz-Zentrum für Polar- und Meeresforschung and Australian Academy of Science DAAD German-Australian mobility grant and BMBF PT/DLR IB FKZ 01DR14006.

LITERATURE CITED

Bachvaroff, T. R., Williams, E., Jagus, R. & Place, A. R. (2015) A noncryptic noncanonical multi-module PKS/NRPS found in dinoflagellates. *The 16th International Conference on Harmful Algae.* Wellington, New Zealand.

Bayer, T., Aranda, M., Sunagawa, S., Yum, L. K., DeSalvo, M. K., Lindquist, E., Coffroth, M. A., Voolstra, C. R. & Medina, M. 2012. *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS ONE,* 7:e35269.

Beedessee, G., Hisata, K., Roy, M. C., Satoh, N. & Shoguchi, E. 2015. Multifunctional polyketide synthase genes identified by genomic survey of the symbiotic dinoflagellate, *Symbiodinium minutum*. *BMC Genomics,* 16:1.

Brown, A.P., Slabas, A.R. & Rafferty, J.B. 2010. Fatty acid biosynthsis in plants-metabolic pathways, structure and organization. In: Wada, H., Murata, N. 9eds), *Lipids in Photosynthesis*. Springer. Dordrecht, The Netherlands, 11-34

Cane, D. E., Walsh, C. T. & Khosla, C. 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science,* 282:63-8.

Chinain, M., Faust, M. A. & Pauillac, S. 1999. Morphology and molecular analyses of three toxic species of *Gambierdiscus* (Dinophyceae): *G. pacificus*, sp. nov., *G. australes*, sp. nov., and *G. polynesiensis*, sp. nov. *J. Phycol.,* 35:1282-1296.

Chinain, M., Darius, H. T., Ung, A., Cruchet, P., Wang, Z., Ponton, D., Laurent, D. & Pauillac, S. 2010. Growth and toxin production in the ciguatera-causing dinoflagellate *Gambierdiscus polynesiensis* (Dinophyceae) in culture. *Toxicon,* 56:739-750.

Chou, H. N. & Shimizu, Y. 1987. Biosynthesis of brevetoxins. Evidence for the mixed origin of the backbone carbon chain and possible involvement of dicarboxylic acids. *J. Am. Chem. Soc.,* 109:2184-2185.

Eichholz, K., Beszteri, B. & John, U. 2012. Putative monofunctional type I polyketide synthase units: a dinoflagellate-specific feature? *PLoS ONE,* 7:e48624.

Emanuelsson, O., Nielsen, H. & Von Heijne, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.,* 8:978-984.

Faust, M. A. 1995. Observation of sand-dwelling toxic Dinoflagellates (Dinophyceae) from widely differing sites, including two new species. *J. Phycol.,* 31:996-1003.

Finn, R. D., Clements, J. & Eddy, S. R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.,* 39:W29-W37.

Fleming, L. E., Baden, D. G., Bean, J. A., Weisman, R. & Blythe, D. G. 1998. Seafood toxin diseases: issues in epidemiology and community outreach. *In*: Reguera, B., Blanco, J., Fernandez, M. L. & Wyatt, T. (eds.) Harmful Algae. Xunta de Galicia and Intergovernmental Oceanographic Commission of UNESCO.245-248.

Fraga, S. & Rodríguez, F. 2014. Genus *Gambierdiscus* in the Canary Islands (NE Atlantic Ocean) with description of *Gambierdiscus silvae* sp. nov., a new potentially toxic epiphytic benthic dinoflagellate. *Protist,* 165:839-853.

Fraga, S., Rodriguez, F., Caillaud, A., Diogene, J., Raho, N. & Zapata, M. 2011. *Gambierdiscus excentricus* sp. nov. (Dinophyceae), a benthic toxic dinoflagellate from the Canary Islands (NE Atlantic Ocean). *Harmful Algae,* 11:10-22.

Friedman, M. A., Fleming, L. E., Fernandez, M., Bienfang, P., Schrank, K., Dickey, R., Bottein, M.-Y., Backer, L., Ayyar, R. & Weisman, R. 2008. Ciguatera fish poisoning: treatment, prevention and management. *Mar. Drugs,* 6:456-479.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.,* 36:3420-3435.

Guillard, R. R. L. & Ryther, J. H. 1962. Studies of marine planktonic diatoms: I. *Cyclotella nana hustedt*, and *Detonula confervacea* (cleve) gran. *Can. J. Microbiol.,* 8:229-239.

Hernández-Becerril, D. U. & Almazán Becerril, A. 2004. Especies de dinoflagelados del género *Gambierdiscus* (Dinophyceae) del Mar Caribe mexicano. *Rev. Biol. Trop.,* 52:77-87.

Holland, W. C., Litaker, R. W., Tomas, C. R., Kibler, S. R., Place, A. R., Davenport, E. D. & Tester, P. A. 2013. Differences in the toxicity of six *Gambierdiscus* (Dinophyceae) species measured using an in vitro human erythrocyte lysis assay. *Toxicon,* 65:15-33.

Hou, Y. & Lin, S. 2009. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. *PLoS ONE,* 4:e6978.

Huang, Y., Niu, B., Gao, Y., Fu, L. & Weizhong, L. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26:680-682.

Jaeckisch, N., Yang, I., Wohlrab, S., Glöckner, G., Kroymann, J., Vogel, H., Cembella, A. & John, U. 2011. Comparative genomic and transcriptomic characterization of the toxigenic marine dinoflagellate *Alexandrium ostenfeldii*. *PLoS ONE,* 6:e28012.

Jenke-Kodama, H., Sandmann, A., Müller, R. & Dittmann, E. 2005. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.,* 22:2027-2039.

John, U., Beszteri, B., Derelle, E., Van de Peer, Y., Read, B., Moreau, H. & Cembella, A. 2008. Novel insights into evolution of protistan polyketide synthases through phylogenomic analysis. *Protist,* 159:21-30.

Kalaitzis, J. A., Chau, R., Kohli, G. S., Murray, S. A. & Neilan, B. A. 2010. Biosynthesis of toxic naturally-occurring seafood contaminants. *Toxicon,* 56:244-258.

Katoh, K., Misawa, K., Kuma, K. i. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.,* 30:3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28:1647-1649.

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., Armbrust, E. V., Archibald, J. M., Bharti, A. K., Bell, C. J., Beszteri, B., Bidle, K. D., Cameron, C. T., Campbell, L., Caron, D. A., Cattolico, R. A., Collier, J. L., Coyne, K., Davy, S. K., Deschamps, P., Dyhrman, S. T., Edvardsen, B., Gates, R. D., Gobler, C. J., Greenwood, S. J., Guida, S. M., Jacobi, J. L., Jakobsen, K. S., James, E. R., Jenkins, B., John, U., Johnson, M. D., Juhl, A. R., Kamp, A., Katz, L. A., Kiene, R., Kudryavtsev, A., Leander, B. S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A. M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M. A., Murray, S., Nadathur, G., Nagai, S., Ngam, P. B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M. C., Rengefors, K., Romano, G., Rumpho, M. E., Rynearson, T., Schilling, K. B., Schroeder, D. C., Simpson, A. G. B., Slamovits, C. H., Smith, D. R., Smith, G. J., Smith, S. R., Sosik, H. M., Stief, P., Theriot, E., Twary, S. N., Umale, P. E., Vaulot, D., Wawrik, B., Wheeler, G. L., Wilson, W. H., Xu, Y., Zingone, A. & Worden, A. Z. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.,* 12:e1001889.

Khosla, C., Gokhale, R. S., Jacobsen, J. R. & Cane, D. E. 1999. Tolerance and specificity of polyketide synthases. *Annu. Rev. Biochem.,* 68:219-253.

Kimura, K., Okuda, S., Nakayama, K., Shikata, T., Takahashi, F., Yamaguchi, H., Skamoto, S., Yamaguchi, M. & Tomaru, Y. 2015. RNA sequencing revealed numerous polyketide synthase genes in the harmful dinoflagellate *Karenia mikimotoi*. *PLoS ONE,* 10:e0142731.

Kohli, G. S., Farrell, H. & Murray, S. A. 2015a. *Gambierdiscus*, the cause of ciguatera fish poisoning: an increased human health threat influenced by climate change. *In*: Botana, L. M., Louzao, M. C. & Vilarino, N. (eds.) Climate change and marine and freshwater toxins. DE Gruyter, Berlin. 1:271-310.

Kohli, G. S., John, U., Van Dolah, F. M. & Murray, S. A. 2016. Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes. *ISME J.*

Kohli, G. S., John, U., Figueroa, R. I., Rhodes, L. L., Harwood, D. T., Groth, M., Bolch, C. J. S. & Murray, S. A. 2015b. Polyketide synthesis genes associated with toxin production in two species of *Gambierdiscus* (Dinophyceae). *BMC Genomics,* 16:410.

Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. & Turgeon, B. G. 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. USA,* 100:15670-15675.

Leadlay, P. F., Staunton, J., Oliynyk, M., Bisang, C., Cortes, J., Frost, E., Hughes-Thomas, Z. A., Jones, M. A., Kendrew, S. G., Lester, J. B., Long, P. F., McArthur, H. A., McCormick, E. L., Oliynyk, Z., Stark, C. B. & Wilkinson, C. J. 2001. Engineering of complex polyketide biosynthesis--insights from sequencing of the monensin biosynthetic gene cluster. *J. Ind. Microbiol. Biotechnol.,* 27:360-7.

Leaw, C.-P., Lim, P.-T., Tan, T.-H., Tuan-Halim, T. N., Cheng, K.-W., Ng, B.-K. & Usup, G. 2011. First report of the benthic dinoflagellate, *Gambierdiscus belizeanus* (Gonyaulacales: Dinophyceae) for the east coast of Sabah, Malaysian Borneo. *Phycol. Res.,* 59:143-146.

Lee, M. S., Qin, G., Nakanishi, K. & Zagorski, M. G. 1989. Biosynthetic studies of brevetoxins, potent neurotoxins produced by the dinoflagellate *Gymnodinium breve*. *J. Am. Chem. Soc.,* 111:6234-6241.

Lee, M. S., Repeta, D. J., Nakanishi, K. & Zagorski, M. G. 1986. Biosynthetic origins and assignments of carbon 13 NMR peaks of brevetoxin B. *J. Am. Chem. Soc.,* 108:7855-7856.

Lewis, R. J. & Holmes, M. J. 1993. Origin and transfer of toxins involved in ciguatera. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology,* 106:615-628.

Lewis, R. J., Sellin, M., Poli, M. A., Norton, R. S., MacLeod, J. K. & Sheil, M. M. 1991. Purification and characterization of ciguatoxins from moray eel (*Lycodontis javanicus, Muraenidae*). *Toxicon,* 29:1115-27.

Lidie, K. B. & Van Dolah, F. M. 2007. Spliced leader RNA‐mediated trans‐splicing in a dinoflagellate, *Karenia brevis*. *J. Eukaryot. Microbiol.,* 54:427-435.

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., Ji, Z., Cai, M., Zhuang, Y., Shi, X., Lin, L., Wang, L., Wang, Z., Liu, X., Yu, S., Zeng, P., Hao, H., Zou, Q., Chen, C., Li, Y., Wang, Y., Xu, C., Meng, S., Xu, X., Wang, J., Yang, H., Campbell, D. A., Sturm, N. R., Dagenais-

Bellefeuille, S. & Morse, D. 2015. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science,* 350:691-694.

Litaker, R. W., Vandersea, M. W., Faust, M. A., Kibler, S. R., Chinain, M., Holmes, M. J., Holland, W. C. & Tester, P. A. 2009. Taxonomy of *Gambierdiscus* including four new species, *Gambierdiscus caribaeus*, *Gambierdiscus carolinianus, Gambierdiscus carpenteri* and *Gambierdiscus ruetzleri* (Gonyaulacales, Dinophyceae). *Phycologia,* 48:344-390.

Litaker, R. W., Vandersea, M. W., Faust, M. A., Kibler, S. R., Nau, A. W., Holland, W. C., Chinain, M., Holmes, M. J. & Tester, P. A. 2010. Global distribution of ciguatera causing dinoflagellates in the genus *Gambierdiscus*. *Toxicon,* 56:711-730.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C. & Bryant, S. H. 2014. CDD: NCBI's conserved domain database. *Nucleic Acids Res.*

Meyer, J. M., Rödelsperger, C., Eichholz, K., Tillmann, U., Cembella, A., McGaughran, A. & John, U. 2015. Transcriptomic characterisation and genomic glimps into the toxigenic dinoflagellate *Azadinium spinosum*, with emphasis on polykeitde synthase genes. *BMC Genomics,* 16:27.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.-Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J. A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D. & Finn, R. D. 2014. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.*

Monroe, E. A. & Van Dolah, F. M. 2008. The toxic dinoflagellate *Karenia brevis* encodes novel type I-like polyketide synthases containing discrete catalytic domains. *Protist,* 159:471-482.

Munir, S., Siddiqui, P. J. A. & Morton, S. L. 2011. The occurrence of the ciguatera fish poisioning producing dinoflagellate genus *Gambieridiscis* in Pakistan waters. *Algae,* 26:371-325.

Murata, M., Legrand, A. M., Ishibashi, Y., Fukui, M. & Yasumoto, T. 1990. Structures and configurations of ciguatoxin from the moray eel *Gymnothorax javanicus* and its likely precursor from the dinoflagellate *Gambierdiscus toxicus*. *J. Am. Chem. Soc.,* 112:4380-6.

Murata, M., Izumikawa, M., Tachibana, K., Fujita, T. & Naoki, H. 1998. Labeling Pattern of Okadaic Acid from 18O2 and [18O2]Acetate Elucidated by Collision-Induced Dissociation Tandem Mass Spectrometry. *J. Am. Chem. Soc.,* 120:147-151.

Murray, S. A., Diwan, R., Orr, R. J. S., Kohli, G. S. & John, U. 2015. Gene duplication, loss and selection in the evolution of saxitoxin biosynthesis in alveolates. *Mol. Phylogen. Evol.,* 92:165-180.

Nagai, H., Murata, M., Torigoe, K., Satake, M. & Yasumoto, T. 1992. Gambieric acids, new potent antifungal substances with unprecedented polyether structures from a marine dinoflagellate *Gambierdiscus toxicus*. *J. Org. Chem.,* 57:5448-5453.

Nascimento, S. M., Melo, G., Salgueiro, F., Diniz, B. d. S. & Fraga, S. 2015. Morphology of *Gambierdiscus excentricus* (Dinophyceae) with emphasis on sulcal plates. *Phycologia,* 54:628-639.

Nishimura, T., Sato, S., Tawong, W., Sakanari, H., Uehara, K., Shah, M. M. R., Suda, S., Yasumoto, T., Taira, Y., Yamaguchi, H. & Adachi, M. 2013. Genetic Diversity and Distribution of the Ciguatera-Causing Dinoflagellate *Gambierdiscus* spp. (Dinophyceae) in Coastal Areas of Japan. *PLoS ONE,* 8:e60882.

Oliynyk, M., Stark, C. B. W., Bhatt, A., Jones, M. A., Hughes‐Thomas, Z. A., Wilkinson, C., Oliynyk, Z., Demydchuk, Y., Staunton, J. & Leadlay, P. F. 2003. Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in *Streptomyces cinnamonensis* and evidence for the role of monB and monC genes in oxidative cyclization. *Mol. Microbiol.,* 49:1179-1190.

Orr, R. J. S., Murray, S. A., Stüken, A., Rhodes, L. & Jakobsen, K. S. 2012. When naked became armored: An eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS ONE,* 7:e50004.

Parra, G., Bradnam, K. & Korf, I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics,* 23:1061-1067.

Pawlowiez, R., Morey, J. S., Darius, H. T., Chinain, M. & Van Dolah, F. M. 2014. Transcriptome sequencing reveals single domain Type I-like polyketide synthases in the toxic dinoflagellate *Gambierdiscus polynesiensis*. *Harmful Algae,* 36:29-37.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., Bateman, A. & Finn, R. D. 2012. The Pfam protein families database. *Nucleic Acids Res.,* 40:D290-D301.

Rein, K. S. & Snyder, R. V. 2006. The biosynthesis of polyketide metabolites by dinoflagellates. *Adv. Appl. Microbiol.,* 59:93-125.

Rhodes, L., Harwood, T., Smith, K., Argyle, P. & Munday, R. 2014. Production of ciguatoxin and maitotoxin by strains of *Gambierdiscus australes*, *G. pacificus* and *G. polynesiensis* (Dinophyceae) isolated from Rarotonga, Cook Islands. *Harmful Algae,* 39:185-190.

Rhodes, L. L., Smith, K. F., Munday, R., Selwood, A. I., McNabb, P. S., Holland, P. T. & Bottein, M.-Y. 2010. Toxic dinoflagellates (Dinophyceae) from Rarotonga, Cook Islands. *Toxicon,* 56:751-758.

Roeder, K., Erler, K., Kibler, S., Tester, P., Van The, H., Nguyen-Ngoc, L., Gerdts, G. & Luckas, B. 2010. Characteristic profiles of Ciguatera toxins in different strains of *Gambierdiscus* spp. *Toxicon,* 56:731-738.

Roy, S. & Morse, D. 2012. A Full Suite of Histone and Histone Modifying Genes Are Transcribed in the Dinoflagellate *Lingulodinium*. *PLoS ONE,* 7:e34340.

Ryan, D., Pepper, A. & Campbell, L. 2014. De novo assembly and characterization of the transcriptome of the toxic dinoflagellate *Karenia brevis*. *BMC Genomics,* 15:888.

Salcedo, T., Upadhyay, R. J., Nagasaki, K. & Bhattacharya, D. 2012. Dozens of toxin-related genes are expressed in a nontoxic strain of the dinoflagellate *Heterocapsa circularisquama*. *Mol. Biol. Evol.,* 29:1503-1506.

Satake, M., Murata, M. & Yasumoto, T. 1993. Gambierol: a new toxic polyether compound isolated from the marine dinoflagellate *Gambierdiscus toxicus*. *J. Am. Chem. Soc.,* 115:361-362.

Shimizu, Y. 2003. Microalgal metabolites. *Curr. Opin. Microbiol.,* 6:236-43.

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., Toyoda, A., Kuroki, Y., Fujiyama, A., Medina, M., Coffroth, Mary A., Bhattacharya, D. & Satoh, N. 2013. Draft Assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.,* 23:1399-1408.

Skinner, M. P., Brewer, T. D., Johnstone, R., Fleming, L. E. & Lewis, R. J. 2011. Ciguatera Fish Poisoning in the Pacific Islands (1998 to 2008). *PLoS Neglected tropical diseases,* 5:e1416.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* 22:2688-2690.

Stüken, A., Orr, R. J. S., Kellmann, R., Murray, S. A., Neilan, B. A. & Jakobsen, K. S. 2011. Discovery of nuclear-encoded genes for the neurotoxin saxitoxin in dinoflagellates. *PLoS ONE,* 6:e20096.

Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. & Deng, Z. 2003. A Complete Gene Cluster from *Streptomyces nanchangensis* NS3226 Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin. *Chem. Biol.,* 10:431-441.

The, H. V. 2009. Sinh Hoc Tao Hai Roi Co Vo Song Day Vung Bien Ven Bo Viet Nam. Institute of Oceanography, Vinh Nguyen, Nha Trang, Nha Trang.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.,* 22:4673-4680.

Veldhuis, M. J. W., Cucci, T. L. & Sieracki, M. E. 1997. Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological Implications. *J. Phycol.,* 33:527-541.

Watanabe, R., Uchida, H., Suzuki, T., Matsushima, R., Nagae, M., Toyohara, Y., Satake, M., Oshima, Y., Inoue, A. & Yasumoto, T. 2013. Gambieroxide, a novel epoxy polyether compound from the dinoflagellate *Gambierdiscus toxicus* GTP2 strain. *Tetrahedron,* 69:10299-10303.

White, S.W., Zheng, J., Zhang, Y.M. & Rock, C.O. 2005. The structural biology of type II fatty acid biosynthesis. *Annu. Rev. Biochem*., 74:791-831.

Wright, J. L. C., Hu, T., McLachlan, J. L., Needham, J. & Walter, J. A. 1996. Biosynthesis of DTX-4: confirmation of a polyketide pathway, proof of a baeyer−villiger oxidation step, and evidence for an unusual carbon deletion process. *J. Am. Chem. Soc.,* 118:8757-8758.

Yasumoto, T., Igarashi, T., Legrand, A. M., Cruchet, P., Chinain, M., Fujita, T. & Naoki, H. 2000. Structural elucidation of ciguatoxin congeners by fast-atom bombardment tandem mass spectroscopy. *J. Am. Chem. Soc.,* 122:4988-4989.

FIGURE LEGENDS

Figure 1: Phylogenetic analysis showing a clear distinction between type II 3-ketoacyl ACP synthase II and type I ketosynthase domains of polyketide synthases and fatty acid synthases. The alignment consisted of 531 characters and 103 representative sequences from dinoflagellates, apicomplexa, chromera, *Vitrella*, stramenopiles, haptophytes, green algae, plants, fungi, animal and bacteria showing the position of each major group in the phylogenetic tree. Analysis was inferred in RAxML using GAMMA model of rate heterogeneity and 1000 bootstraps.

Figure 2: Concatenated phylogeny of five enzymes involved in type II fatty acid synthesis (3-ketoacyl ACP synthase III, s-malonyltransacylase, trans3-ketoacyl ACP reductase, 3-hydroxyacyl-ACP dehydratase and enoyl-ACP reductase) from 18 dinoflagellates and one other alveolate *Chromera velia* which was used as an outgroup. The alignment consisted of 1431 characters and the phylogeny was inferred using RAxML, GAMMA model of rate heterogeneity and 1000 bootstraps.

Figure 3: Phylogenetic analysis of ketoacyl synthase (KS) domains from prokaryotic and eukaryotic type I & II polyketide synthases (PKS) and fatty acid synthases. The alignment consisted of 573 characters, 478 sequences encoding KS domain from 10 dinoflagellates and 30 other prokaryotic and eukaryotic taxa. Analysis was inferred using RAxML, GAMMA model of rate heterogeneity and 1000 bootstraps.

Figure 4: A schematic representation of type I polyketide synthase complex encoded by contig_50464 in *G. polynesiensis* transcriptome. Biosynthetic pathway and a structure of a compound resembling partial carbon backbone of a polyether ladder that could be produced by the megansynthase has been proposed. Following abbreviations were used: KS: ketosynthase; KR: ketoreductase; ACP: acyl carrier protein; AT: acyl transferase; DH: dehydratase; ER: enoyl reductase; TE: thioestrase.

Figure 5: Proposed mechanism of Pacific-ciguatoxin (P-CTX) production. Polyene (preciguatoxin) produced by polyketide biosynthesis undergoes epoxidation and epoxide cyclisation to form ciguatoxin.

SUPPORTING INFORMATION

Figure S1: Phylogenetic analysis of 18s ribosomal rDNA from various *Gambierdiscus*, *Fukuyoa* and *Alexandrium* species. The alignment consisted of 1822 characters. Analysis was inferred using RAxML, GTRCAT model of rate heterogeneity and 1000 bootstraps.

Table S1: A list of all the sequences encoding the histone proteins (H2A, H2B, H3, H4) found in the gene catalogue of *G. excentrucus* and *G. polynesiensis*.

Table S2: Description of sequences from *Gambierdiscus excentricus* and *Gambierdiscus polynesiensis* encoding essential enzymes for various metabolic pathways (*G. excentricus*: 159 out of 167 and *G. polynesiensis*: 161 out of 167 enzymes were present).

Table S3: A list of type II FAS genes found in *Gambierdiscus australes*, *G. belizeanus*, *G. excentricus* and *G. polynesiensis*.

Table S4: Polyketide synthase genes found in *Gambierdiscus australes*, *G. belizeanus*, *G. excentricus* and *G. polynesiensis* transcriptomes. Sequences names with a prefix "sl" and/or "polyA" denote the presence of Spliced leader and polyA tails at the 5' and 3' end of the sequences respectively.

Table S5: Sequence properties of the transcripts encoding epoxidases and epoxide hydrolases identified in *G. polynesiensis* and *G. excentricus*.

Supplementary file 1: Alignments used to develop figures 1, 2, 3 and HMM databases for various enzymes.

Table 1: Geographic distribution and toxicity of *Gambierdiscus* species used in this study.

Species	Geographical distribution	Toxicity			
		Various assays		LC-MS	
		CTX	MTX	CTX	MTX
G. polynesiensis	French Polynesia (Chinain et al. 1999), Pakistan (Munir et al. 2011), Nha Trang-Vietnam (The 2009), Cook Islands (Rhodes et al. 2014)	MBA-positive (Chinain et al. 1999, Rhodes et al. 2014), RBA- positive (Chinain et al. 2010)	MBA-positive (Chinain et al. 1999, Rhodes et al. 2014)	Yes (Chinain et al. 2010, Rhodes et al. 2014)	MTX3 only detected (Rhodes et al. 2014)
G. excentricus	Canary Islands (Fraga et al. 2011), Brazil (Nascimento et al. 2015), Oman (Saburova pers. com in Nascimento et al. 2015)	NCBA-positive (Fraga et al. 2011)	NCBA-positive (Fraga et al. 2011)	N/K	N/K
G. australes	French Polynesia (Chinain et al. 1999), Japan (Nishimura et al. 2013), Cook Islands (Rhodes et al. 2010), Hawaii USA (Litaker et al. 2009), Pakistan (Munir et al. 2011), Canary Islands (Fraga and Rodríguez 2014)	MBA-positive (Chinain et al. 1999, Nishimura et al. 2013, Rhodes et al. 2010), RBA-positive (Chinain et al. 2010)	HELA-positive , MBA-positive (Chinain et al. 1999, Nishimura et al. 2013, Rhodes et al. 2010)	N/D (Rhodes et al. 2010, Rhodes et al. 2014, Kohli et al. 2015b)	MTX and MTX3 detected (Rhodes et al. 2014)
G. belizeanus	Belize (Faust 1995), Florida (Litaker et al. 2009) , Mexican Caribbean (Hernández-Becerril and Almazán Becerril 2004), Malaysia (Leaw et al. 2011), Pakistan (Munir et al. 2011), Queensland, Australia (murray unpubl. Data), St. Barthelemy Island-Caribbean (Litaker et al. 2010)	RBA- positive (Chinain et al. 2010)	HELA-positive (Holland et al. 2013)	N/D (Kohli et al. 2015b)	MTX3 only detected (Kohli et al. 2015b)

Table 2: Coverage and annotation statistics of *G. excentricus* and *G. polynesiensis* gene catalogue. An e-value cut-off of 10^{-9} was applied during BLASTx analysis.

Coverage	**No. of contigs**	**Length (mean)**	**BLASTx analysis**			**PKS sequences**
			Annotated match	**Non-Annotated match**	**No match**	
Gambierdiscus excentricus						
1x-5x	4109	300-1596 (415.1)	343	273	3493	
5x-20x	17443	300-7742 (814.2)	2736	2358	12349	12
20x-50x	15779	300-7378 (1393.9)	4058	2896	8825	28
50x-100x	14409	300-7013 (1441)	4431	2865	7113	38
100x-1000x	25122	300-7279 (1182.8)	7395	4937	12790	36
1000x-10000x	487	300-5417 (885.1)	146	90	251	
> 10,000x	43	311-1756 (807.2)	22	10	11	
16x*	1*	245*	-	1*	-	1*
Total number of contigs (percentage)	77393	245-7742 (1148)	19131 (24.7%)	13430 (17.4%)	44832 (57.9%)	115
Gambierdiscus polynesiensis						
1x-5x	1105	300-722 (355.2)	41	57	1007	
5x-20x	12367	300-8556 (505.9)	1064	938	10365	
20x-50x	8363	300-29981 (811.1)	931	949	6483	5
50x-100x	8902	300-30289 (962.7)	1304	1325	6273	10
100x-1000x	57449	300-15192 (1017.4)	13042	9780	34627	101
1000x-10000x	26712	300-7993 (972.6)	7484	5488	13740	46
> 10,000x	872	300-6418 (695.8)	326	157	389	
69x-993x*	10*	200-295 (245.8)*	-	10*	-	10*
Total number of contigs (percentage)	115780	200-30289 (924.5)	24192 (20.9%)	18704 (16.1%)	72884 (63%)	172

* PKS sequences below 300bp

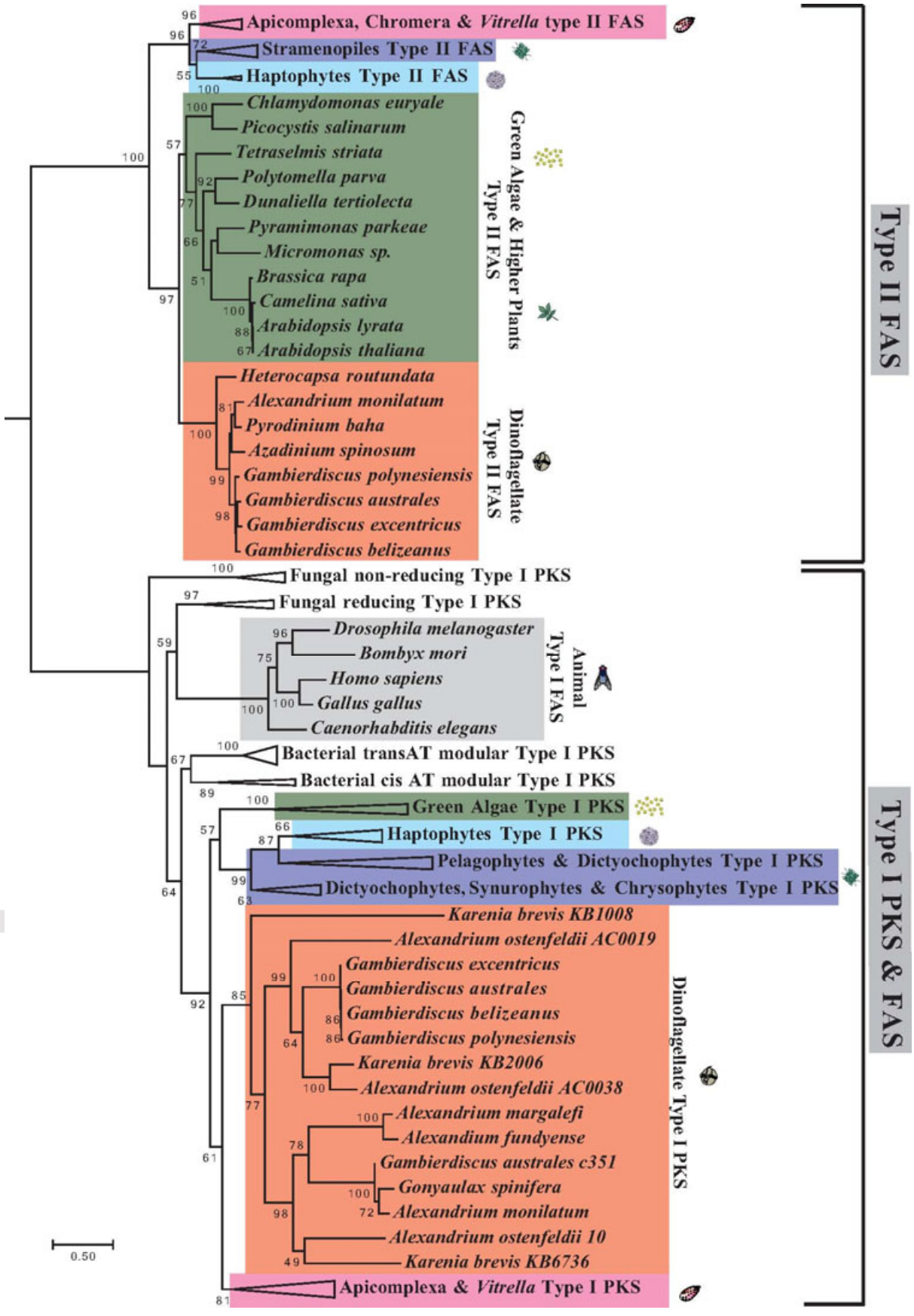
Table 3: Total number of polyketide synthase (PKS) associated domains found in four *Gambierdiscus* transcriptomes. Results for *Gambierdiscus australes* and *G. belizeanus* were taken from Kohli et al., 2015. Following abbreviations were used: KS: ketosynthase; KR: ketoreductase; ACP: acyl carrier protein; AT: acyl transferase; DH: dehydratase; ER: enoyl reductase; TE: thioestrase; A: Adenylation domain of non-ribosomal peptide synthase.

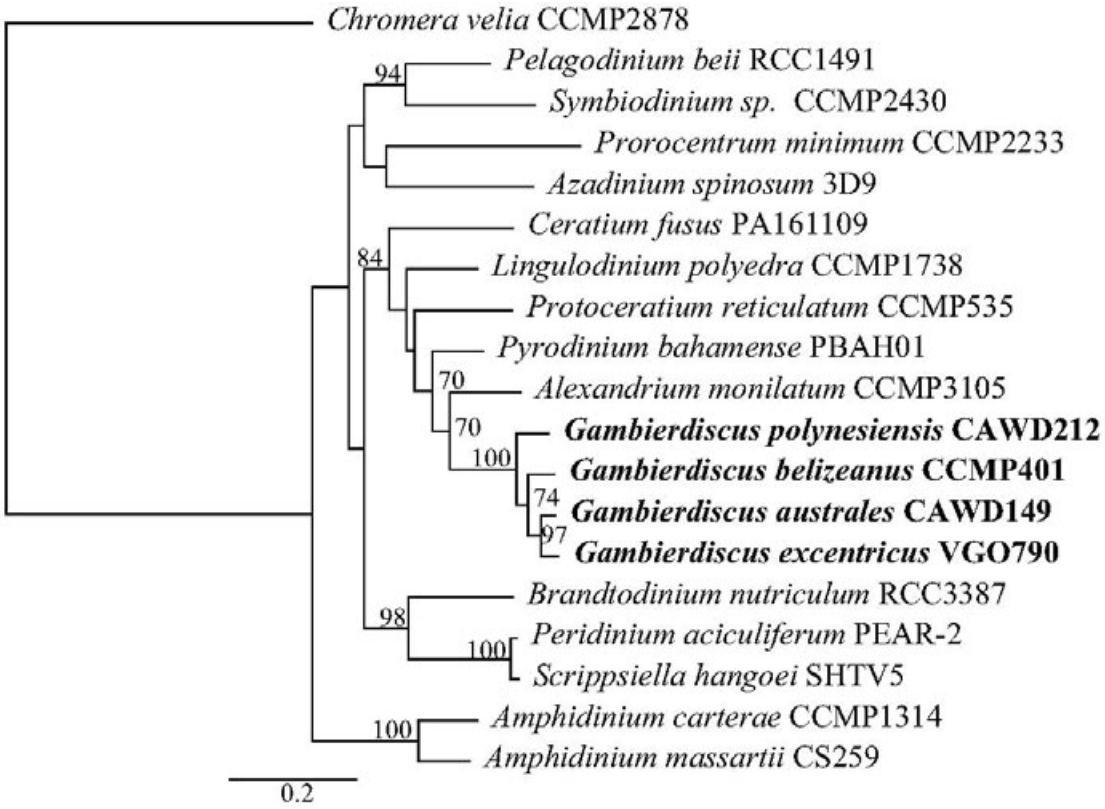
Sequences with single PKS domains			
Organism	**Total number of Transcripts (Full/Partial) domain**	**Encoding Domain**	**Average GC content (%) (Full/Partial) domain**
G. excentricus	86/20	KS	60.4/62.7
	7/0	KR	59.8
G. polynesiensis	73/70	KS	59.9/60.9
	7/1	KR	58.5/66.8
G. australes	90/12	KS	60.4/59.1
	7/4	KR	58.8/65.8
G. belizeanus	74/40	KS	60.0/60.3
	6/2	KR	58.6/62.2
Sequences with multiple PKS domains			
Organism	**Sequence Name**	**Encoding Domains**	**GC content (%)**
G. australes	10913	KS-KR-DH-ER-ACP	64.3
	20703	ACP-KS-AT-DH-ER-KR-ACP-TE[#]	59.0
	24973	KS-ACP	58.5
	38797	ACP-KS-AT-KR-ACP	65.1
	45392	KS(Partial)-DH-KR-ER-ACP-TE(Partial)	63.1
	47701	KS-AT	65.9
	64398	ACP-KS(Partial)	63.7
G. belizeanus	105365	ACP-KS(Partial)	67.6
	45222	KS-(Partial)-AT-TE-KR	66.9
	94589	ACP-KS	63.2
	13526	KS-ACP	61.1
	11718	AT(Partial)-DH-ER-KR-ACP-TE	60.1

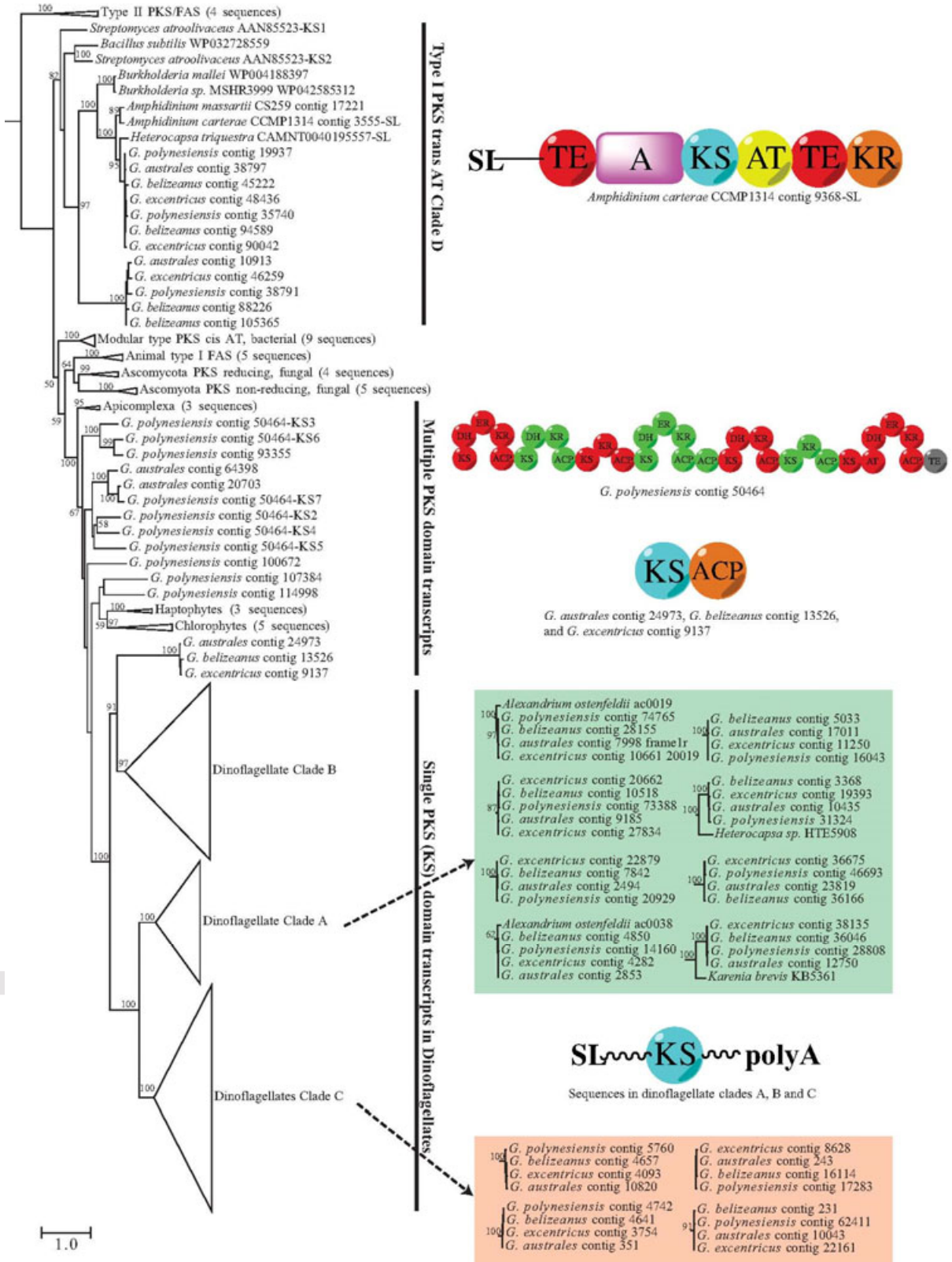
G. excentricus	48436	KS(Partial)-AT-TE-KR	68.2
	9137	KS-ACP	60.3
	6057	DH-ER-KR-ACP-TE	60.4
G. polynesiensis	1250	KS(Partial)-AT-DH-KR	58.4
	19937	TE-A-ACP-KS(Partial)*	64.4
	20287	(Partial)KS-DH-KR(Partial)	65.4
	21774	PP-KS(Partial)	63.8
	35740	KS(Partial)-AT-TE-KR(Partial)	67.4
	38791	KS(Partial)-KR-DH-ER-PP	65.5
	50464	Partial-KS(1)-DH-ER-KR-ACP-KS(2)-DH-KR-ACP-KS(3)-KR-ACP-KS(4)-DH-ER-KR-ACP-ACP-KS(5)-DH-KR-ACP-KS(6)-KR-ACP-KS(7)-AT-DH-ER-KR-ACP-TE	57.9%
	64449	ACP-KS(Partial)	64.3
	88992	ACP-KS(Partial)	64.0
	93355	DH-KR-ACP-KS-KR	65.4
	100672	KR-ACP-KS	63.4
	107384	KR-ACP-KS	68.3
	114998	ACP-KS_DH	66.8
	117253	KR(Partial)-ACP-KS(Partial)	68.1
	118767	A-ACP-KS(Partial)*	65.7
	135859	ER-KR-ACP-KS(Partial)	55.9
	136511	(Partial)KS-DH(Partial)	65.4
	138240	(Partial)KS-KR-ACP-KS(Partial)	59.2
	139846	ACP-KS(Partial)	55.5
	141304	(Partial)KS-AT(Partial)	56.2
	144687	KR-ACP-KS(Partial)	55.5

*Hybrid polyketide synthase non-ribosomal peptide synthases

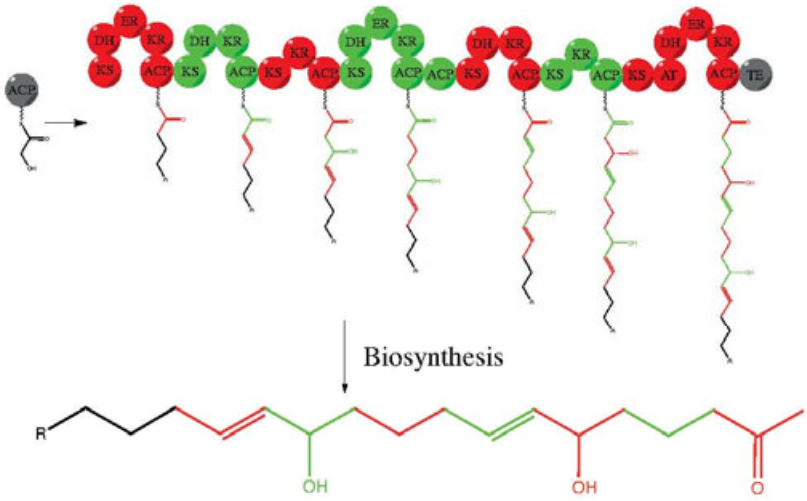
#Sequences similar *to G. polynesiensis* contig 50464







Type I modular Polyketide synthase complex



Carbon backbone of putative polyether ladder compound